Sampling Assumptions and the Size Principle in Property Induction

Philip M. Fernbach (philip_fernbach@brown.edu)
Brown University Department of Cognitive and Linguistic Sciences, Box 1978
Providence, RI 02912 USA

Abstract

The 'size principle' emphasized in recent Bayesian models of inductive generalization (Kemp & Tenenbaum, 2003; Sanjana & Tenenbaum, 2003; Tenenbaum & Griffiths, 2001) is tested in the domain of property induction. As predicted by the model, the size principle is obeyed more frequently when a strong sampling assumption is explicit than when sampling is weak or unspecified, but it is not followed consistently. This implies that, although people are sometimes sensitive to sampling assumptions as specified by the Bayesian framework, models based on a strong sampling assumption may not provide general accounts of property induction.

Introduction

From the time of their introduction in the 16th century until general acceptance in the 18th century, tomatoes were not widely eaten in Mediterranean Europe. Due to their morphological resemblance to the nightshade plant, they were assumed poisonous and avoided. This is a striking — if unsuccessful — example of a common strategy that human beings utilize to make sense of the world: to induce possession of a property by one category from another. Human beings perform these inductions naturally, often from few examples and even between categories that are prima facie dissimilar. Understanding the nature of this type of inference is a major project of cognitive science.

Empirical studies of property induction have been conducted since the mid 1970s using argument strength ratings for premise-conclusion pairs and blank predicates. Figure 1 provides an example.

	Salmon have sesamoid bones
	Lions have sesamoid bones
Therefore	Polar Bears have sesamoid bones

Figure 1: Property induction argument

A host of phenomena have been observed concerning how people evaluate the strength of such arguments (reviewed by Sloman & Lagnado, 2005). Several models have been proposed to account for the phenomena. The models fall into two camps and as such are revealing of a deeper methodological divergence in cognitive science. On the one hand are models directly motivated by empirical work and aimed at the level of algorithm (Marr, 1982). The two best known were proposed in the 1990s (Osherson et al., 1990; Sloman, 1993) and rely on notions of similarity and feature matching. These models tend to be good explanatory accounts of the phenomena that have been studied but their scope is limited.

On the other hand are models which come from the tradition of rational analysis (Anderson, 1990). These models seek to identify the computational problem the system is trying to solve and propose an optimal solution. A common theme of the models of this type has been that they utilize the tools of Bayesian statistics. Heit (1998) was the first to apply the Bayesian framework to property induction, providing an account of several of the empirical phenomena. More recently a family of models has emerged that has advanced the project by introducing a principled likelihood calculation based on sampling assumptions, proposing a general method for generating a hypothesis space and attempting to ground the prior distribution in domain specific theories (Kemp & Tenenbaum, 2003; Tenenbaum & Griffiths, 2001; Sanjana & Tenenbaum, 2003). I refer to this family of models collectively as the Sampling Sensitive Bayesian (SSB) model.

In this paper I evaluate the approaches by testing a key prediction of the SSB model. The SSB model relies on a likelihood calculation that is dependent on assumptions about how the data are sampled. This implies that argument strength judgments should be sensitive to variations in sampling procedures. Specifically, the SSB model generally assumes that human property induction is best described using strong sampling (Kemp & Tenenbaum, 2003; Sanjana & Tenenbaum, 2000) an assumption that examples are drawn at random from the set of objects to which the predicate applies. This assumption leads to a 'size principle,' a preference for smaller hypotheses over larger ones given data that is consistent with both. If the size principle holds, then given more examples within a range, participants should be less willing to generalize outside of that range, a tendency that should manifest itself as an inductive non-monotonicity whereby adding similar premises decreases property induction to a dissimilar conclusion. This phenomenon is not predicted by either Osherson et al's (1990) Similarity/Coverage (SC) model or Sloman's (1993) Feature-Based (FB) model as neither of these models is sensitive to variations in sampling procedures.

To identify the source of this divergence I first present the similarity and feature-based models. I then present the SSB model and show how the size principle emerges from the model and why it does not apply to other models. Lastly I present an argument preference experiment designed at testing for the size principle and discuss the implications of the results.

Similarity and Feature-Based Models

Similarity/Coverage The SC model is based on overall similarity judgments and assumes static category structure. Argument strength is assessed based on three factors. The first is the maximum similarity between the set of premise objects and the conclusion. The second is the similarity of the premise set to all of the members of the immediately super-ordinate category that contains both the premise and the conclusion, referred to as coverage. The third is a free parameter denoting the relative weight of similarity and coverage for an individual subject. Thus, an argument is considered strong to the extent that the premise objects are similar to the conclusion and the premise objects cover the super-ordinate category. The model accounts for a wide variety of empirical phenomena and provides a strong match to human data (Osherson et al, 1990).

Feature-Based The FB model (Sloman, 1993) is expressed as an artificial neural network where input units represent values over a set of features and the output unit represents the presence or absence of some property. The activation of the output unit is determined by two things: the feature overlap of the premise and conclusion and the number (in the binary case) or the 'magnitude' (more generally) of salient features possessed by the conclusion. Thus an argument is considered strong if there is a great deal of overlap of the features of the premise and conclusion and the conclusion has few additional features. The FB model is not based on a notion of similarity. To the extent that similarity is based on feature overlap the model's feature matching rule may approximate similarity, but the basis for generalization is featural, not similarity based. This is an important distinction from the SC model which gives computational primacy to the similarity calculation. Nonetheless, the empirical predictions of the FB model are similar to the SC model. One important exception is that the FB model never predicts non-monotonicity, that adding premises decreases argument strength, despite experimental evidence for it.

Sampling Sensitive Bayesian Model

The SSB model recasts the computational problem in statistical terms: Given data in the form of one or more premise objects that are examples of some concept, what is the probability that the conclusion object also belongs to that concept? The concept corresponds to the full set of objects to which the predicate applies (e.g. animals that have sesamoid bones).

The solution is offered by Bayes' rule which stipulates how to update hypotheses in the light of data. The first step is to identify a hypothesis space, a set of groupings of objects which could conceivably correspond to the concept. A prior distribution is assigned to the hypotheses which denotes how likely each hypothesis is *a priori*. Next, using Bayes' rule, the probability distribution is updated to take into account the data. Finally the status of the conclusion object with respect to the concept is calculated by adding up the posterior probabilities of the hypotheses to which the conclusion belongs. Thus, an argument is considered strong to the extent that the conclusion is a member of hypotheses that have high posterior probability of corresponding to the concept. Note that unlike the SC and FB models, generalization is not based on comparing the premise to the conclusion directly, but rather is mediated by the concept corresponding to the true set of objects to which the predicate applies.

Hypothesis Space and Prior The goal in specifying a hypothesis space is to identify those groupings of objects that humans would consider candidates for correspondence with the concept. This is a difficult task since objects are grouped differently based on domain. Sanjana and Tenenbaum (2003) propose a similarity-based hierarchical clustering approach where clusters and unions of clusters supply the hypothesis space. This is a reasonable and computationally tractable solution but it is imperfect since overall similarity ratings may often misrepresent the way objects are grouped depending on the predicate. For instance, a bat might be considered more similar to a bird overall, but biologically more similar to a whale. The source of the prior distribution is also a difficult issue as prior knowledge varies across domains and individuals. Sanjana and Tenenbaum (2003) posit a prior distribution motivated by rational principles such as preference for simplicity and recent work has attempted to ground the prior in domain-specific theories (Kemp & Tenenbaum, 2003). Resolving these difficulties remains a major task for proponents of the Bayesian framework.

Calculating the Posterior Distribution Bayes' rule specifies that the posterior distribution over hypotheses is proportional to the prior distribution multiplied by the likelihood of the data given each hypothesis

$$p(h|d) \propto p(d|h)p(h) \qquad (1)$$

where $p\ (h \mid d)$ is the posterior probability that a hypothesis (h) corresponds to the set of objects that belong to the concept given some data (d), $p(h)$ is the prior probability assigned to the hypothesis and $p\ (d \mid h)$ is the likelihood of observing the data under that hypothesis.

Thus, given a hypothesis space and a prior distribution, all that is needed to calculate the posterior distribution is the likelihood, the probability distribution over hypotheses of observing the data given that each hypothesis is true and therefore corresponds to the concept. Calculation of the likelihood depends on how the data are sampled. Two sampling procedures are considered, weak and strong. Weak sampling implies that an example is sampled independently of the concept and then given a label specifying whether or not it belongs to the correspondent set of objects. Given weak sampling and data in the form of positive examples of the concept, the likelihood of observing the data given a hypothesis is 1 if the example belongs to the hypothesis and

0 otherwise and as such is a binary indication of whether the example is consistent with the hypothesis.

$$p(d|h) = \begin{cases} 1 & \text{if } d \in h \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

To clarify why this is so, take the scenario of a scientist trying to identify the group of animals that have property 'X'. To simplify the example suppose that there are twenty types of animals in the world, and one example of each type. All of the animals are vertebrates, but just five are mammals. Also assume that the scientist's hypothesis space consists of just two hypotheses, that all vertebrates have property 'X', and that just mammals have it. To test these hypotheses, he chooses an animal from the set of twenty at random and tests it for the property. The animal that he selects happens to be a lion and he finds that it does indeed have property 'X'. This is an instance of weak sampling. The data consists of an example (lion) and a label ('has property 'X'') and the probability of the data is irrespective of the size of the hypothesis. If the larger hypothesis is true and all vertebrates have property 'X', the probability that the lion would be observed to have property 'X' is 1. Likewise, since lions are mammals, if the smaller hypothesis is correct, the probability that the lion would be observed to have property 'X' is also 1. So given weak sampling and data that is consistent with both hypotheses, the likelihood does not favor one over the other.

Strong sampling implies that the example is sampled at random explicitly from the set of objects to which the predicate applies. This makes the data more informative about the nature of that set since the probability of observing that example is tied to the size of the hypothesis. Specifically, the likelihood is the inverse of the size of the hypothesis if the data is consistent with the hypothesis and 0 otherwise. In the case of more than one example drawn independently, the likelihood becomes the inverse of the size of the hypothesis raised to the number of examples

$$p(d|h) = \begin{cases} \frac{1}{|h|^n} & \text{if } d \in h \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where $|h|$ is the number of objects in the hypothesis and n is the number of independently drawn examples. .

As an example of strong sampling imagine a slightly different scenario. The scientist is assessing the same two hypotheses but this time the animals have already been tested for the property and those possessing it placed together in a room. The scientist is allowed to randomly observe one animal from the room at a time. The first animal that he observes happens to be a lion. This is an example of strong sampling. The example is sampled at random from the set of objects that possess the property. Unlike under weak sampling, the probability of the data is not the same for the two hypotheses. Rather, if the smaller hypothesis were true he would have had a one out of five chance of observing a lion, whereas if the larger were true, only a one out of twenty chance.

Size Principle Given the strong sampling assumption the likelihood of a hypothesis decreases exponentially in proportion to its size as new examples are encountered. As posterior probability is proportional to likelihood, this means that smaller hypotheses are favored over larger ones given data that are consistent with both (assuming that the larger hypothesis is not strongly favored *a priori*). Intuitively, this reflects that given random sampling from the concept, seeing a number of examples consistent with a smaller hypothesis would be a coincidence if the larger hypothesis were true and that the psychological plausibility of this coincidence decreases exponentially as more examples are added that are consistent with the smaller hypothesis.

To see how this bears on induction we return to the strong sampling scenario in which there are twenty animals, five of which are mammals. If the scientist observes several examples and each is a mammal, Bayes' rule indicates that he should give high posterior probability to the smaller hypothesis relative to the larger one. The probability of inducing a property to a conclusion is the sum of the posterior probabilities of the hypotheses of which the conclusion is a member. Since the non-mammals are only members of one hypothesis and it has a low posterior probability, the scientist should be reluctant to generalize the property to the non-mammals. This tendency holds for scenarios with more realistic hypothesis spaces and sets of objects and can be stated more generally as an unwillingness to induce a property to a conclusion outside of a range given examples within that range.

In property induction tasks the size principle should manifest itself as a non-monotonicity whereby adding similar examples decreases the strength of an argument whose conclusion is dissimilar. This phenomenon is not predicted by either the SC or FB models. The SC model predicts non-monotonicity only in the case where adding a premise changes the super-ordinate category which is not the case for premises within a range and a conclusion outside of that range. The FB model never predicts non-monotonicity because adding premises can neither decrease feature overlap nor decrease the magnitude of the conclusion. More generally, the Bayesian model is sensitive to sampling because the sampling procedure determines what can be inferred about the relationship between the hypothesis space and the concept. Since the FB and SC models do not rely on the notion of a hypothesis space nor of a concept, phenomena that stem from variations in sampling assumptions cannot be accommodated by the theories.

Bayesian Model Claims It is helpful to view the assertion of a size principle in the SSB model as amounting to two related claims. The first claim is that judgments of argument strength should be sensitive to sampling procedure. This claim is essential to the Bayesian model as it does not rest on a specific modeling assumption but on the general framework of the model.

The second claim is that a strong sampling assumption is appropriate for describing the property induction task at issue. The status of this claim is controversial. Heit (2001) maintains that a non-monotonicity associated with adding premises is inconsistent with many of the findings in the literature; however, there is some evidence in favor of this phenomenon. Sanjana and Tenenbaum (2003) report a non-monotonic effect consistent with the size principle in a property induction experiment, but the effect is small and their experiment departs from the standard property induction paradigm in several ways. Medin et al. (2003) also predict non-monotonicity with addition of similar premises but without relying explicitly on sampling. Rather, they hypothesize that participants assume that the premises are chosen so as to be informative of the category to which the predicate pertains. Adding premises that share a property will increase the association between that property and the predicate and weaken an argument if the conclusion does not possess the property. This is similar to the notion of strong sampling in that it implies that the experimenter purposefully chose the premises from the subset of objects to which the predicate applies. Medin et al. do report some non-monotonicities from adding premises, but the phenomenon does not hold across all their test items.

Experiment

To evaluate the two claims of the Bayesian model I conducted an experiment aimed at identifying if and when the size principle is manifested in property induction. I asked participants to choose between one-premise and three-premise arguments given different cover stories which implied different sampling procedures.

The first claim was tested by contrasting judgments of groups given either weak sampling or strong sampling instructions. According to the SC and FB models, changing sampling assumptions should not alter judgments of argument strength. The SSB model, however, is sensitive to sampling and predicts non-monotonicities with strong sampling but not with weak sampling. To evaluate the second claim, the third group was given instructions that were vague about the sampling procedure as in the conventional task, enabling insight into people's default assumption.

Method

Participants Participants were 41 Brown University graduate and undergraduate student volunteers assigned randomly to three groups, 14 to the ambiguous condition, 13 to weak sampling, and 14 to strong sampling.

Procedure All participants were given the same 10 scenarios to judge. Four were test items and six were dummy scenarios created randomly to eliminate any demand characteristic. As in Figure 2, each scenario consisted of two arguments, labeled 'A' and 'B', and participants had to choose which was stronger on a scale of one to seven. The arguments consisted of premises that attributed some blank biological property to one, two or three animals and a conclusion attributing that property to raccoons. For the test items one argument contained three premises which were all animals very similar to one another and the other contained just one of those animals.

A.

Lions secrete uric acid crystals

Raccoons secrete uric acid crystals

B.

House Cats secrete uric acid crystals

Lions secrete uric acid crystals

Tigers secrete uric acid crystals

Raccoons secrete uric acid crystals

1 2 3 4 5 6 7

Strongly Prefer A … Strongly Prefer B

Figure 2: Example argument strength scenario

Instructions were varied across groups to reflect different sampling assumptions. One group was given instructions that implied strong sampling, one weak sampling and one group was given ambiguous instructions. For the ambiguous group the instructions replicated the ones used in Osherson et al's (1990) original study and asked participants to rate which argument was stronger assuming that the statements above the lines were facts and those below the lines conclusions that follow from those facts.

For both strong sampling and weak sampling groups, a cover story was used to explain how the facts were generated. The story involved two students each writing a paper about raccoons. Premises described facts the the students had uncovered in their research and the conclusion was the statement about raccoons that they were going to put into the paper. Participants were asked to rate which student was more justified in putting the conclusion into their paper.

The difference between the instructions for the strong sampling and weak sampling groups concerned the way the students conducted their research. In the strong sampling condition, the research was described as follows: "Albert and Bob both used the same book for the research, a book of biological facts. Each section of the book covers some property (e.g. animals with Sesamoid bones) and each page of the section contains a picture of an animal with that property. Albert and Bob got their facts by randomly flipping to pages in the appropriate section. So for example if Albert says that Lions have Sesamoid bones and Wolves have Sesamoid bones, he knows that because he looked in the section on Sesamoid bones and then flipped at random to a page and found a picture of a lion and then flipped at random to another page and found a wolf." In the weak sampling condition, the research was described as follows: "Albert and Bob got their facts by choosing an animal and finding out if that animal had some property (e.g. Sesamoid bones) and then choosing another animal and checking for the property and so on. The number of facts in each scenario represents how many animals they looked at in that scenario. So for example, if Albert says that Lions have Sesamoid bones and Wolves have Sesamoid bones, he knows that because he first looked at lions and found they had Sesamoid bones, and then chose to look at Wolves and found that they too had Sesamoid bones. Please note that in a particular scenario the animals listed are the only ones that were looked at."

Model Predictions The SC and FB models make no assumptions about sampling so make the same predictions regardless of instructions. The SC model predicts that three-premise arguments (option A in Figure 2) should be judged stronger than the one-premise arguments (option B in Figure 2) because the premises were chosen so that the three-premise arguments had higher similarity and coverage than the one-premise arguments. In the case of the FB model, the three-premise arguments should be judged stronger or approximately the same. Adding premises so close within a range may not increase feature overlap measurably, but the one-premise arguments should never be favored. The SSB model predicts that one-premise arguments should be favored in the case of strong sampling due to the size principle. For weak sampling, three-premise arguments should be slightly favored because certain hypotheses, such as the one-premise animal alone (e.g. just lions secrete uric acid crystals), have been eliminated increasing the posterior distribution of hypotheses that include the conclusion. There is no prediction inherent in the Bayesian framework for the ambiguous condition as the model can be accommodated to weak or strong sampling depending on how the instructions are interpreted. Kemp & Tenenbaum (2003) and Sanjana & Tenenbaum (2003) assume strong sampling, but the SSB model does not necessitate that assumption.

Table 1: Prediction whether three-premise arguments or one-premise arguments should be judged stronger for each model across the three conditions.

	Ambiguous	Strong Sampling	Weak Sampling
Similarity/ Coverage	3-Premise	3-Premise	3-Premise
Feature-Based	3-Premise or Equal	3-Premise or Equal	3-Premise or Equal
Bayesian	Unspecified	1-Premise	3-Premise

Results

As predicted by the Bayesian Model, variation of sampling assumptions yielded a statistically significant treatment effect (one way ANOVA; F=5.97, p<.01). The direction of the effect was also in line with the Bayesian model. The strong sampling group displayed a greater preference for one-premise arguments than did the weak sampling or ambiguous groups and this result was significant (Figure 3). Pairwise t-tests indicated a statistically significant difference between the strong sampling group and the weak sampling groups (t=2.68, p<.01), and between the strong sampling and ambiguous groups (t=3.18, p<.01), but no difference between the ambiguous and weak sampling groups.

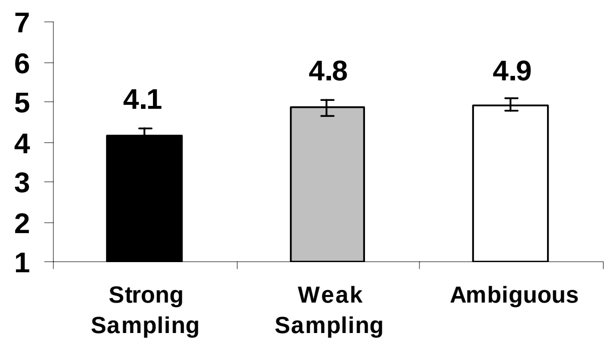

Figure 3: Average preference for three-premise arguments on a 1-7 scale. A score of 7 implies that the three-premise argument was strongly preferred; a score of 1 implies that the one-premise argument was strongly preferred and a score of 4 implies that the one-premise and three-premise arguments were judged equally strong.

Though the strong sampling group showed a greater preference for the one-premise arguments than the weak sampling or ambiguous groups, all groups showed an overall preference for three-premise arguments. In other words, most participants regardless of group failed to display the size principle. Participants in the strong sampling group preferred the one-premise arguments 30% of the time versus 43% for the three-premise arguments. Weak sampling and ambiguous groups behaved similarly to one another preferring the three-premise arguments approximately 73% and 66% of the time respectively (Figure 4).

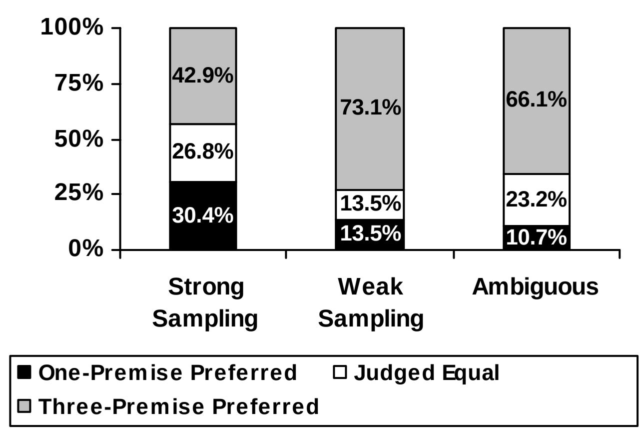

Figure 4: Percentage of scenarios in which three-premise arguments were judged stronger, one-premise arguments were judged stronger and arguments were judged equally strong across all three conditions.

Discussion

The majority of participants across all conditions displayed a monotonic tendency given additional examples within a range and a conclusion outside of that range. The exception was that participants gave a significant percentage of non-monotonic responses in the strong sampling condition. These results indicate that both types of models, the similarity and feature-based and the Bayesian, capture some aspect of property induction since sensitivity to sampling is predicted only by the Bayesian model, whereas a general monotonic tendency fits better with the SC and FB models. Both Heit (1998) and Kemp and Tenenbaum (2003) note an empirical correspondence between the Bayesian model and the similarity and feature-based models. They suggest that the two types of models may be best viewed as complementary rather than competitive. The SC model, for example, may be a heuristic-based approximation of the SSB model as implemented in human beings while the SSB model provides a computational level explanation of why the SC model should work. These results highlight an empirical lack of correspondence and therefore imply a somewhat different account.

There is a tension between positing a computational level model and making empirical predictions. If the algorithm that the system uses only approximates the optimal computation then it does not follow that phenomena predicted by the computational level model should be observed. Of course, the SSB model is only testable to the degree that it is committed to specific predictions. The SSB model sets up the computation as a choice between hypotheses that is sensitive to sampling. While this characterization does provide new insights into the inductive process, people do not fully conform to it. Dual process theory (Evans & Over, 1996; Sloman, 1996; Stanovich & West, 2000) provides a possible solution. The size principle non-monotonicity may be attributable to an effortful rule-based reasoning process, whereas the more common monotonic response is a result of the intuitive process. According to this account, The SC and FB models, or some other relatively simple heuristic, represent an intuitive associative process that occurs effortlessly and sometimes leads to counter-normative results. The SSB model, in contrast, provides the computational level account of how the rule-based system goes about integrating sampling information and other considerations into an evaluation of candidate hypotheses. This would explain both why some people were sensitive to the sampling manipulation and why most people were not on the assumption that only a minority put in the required effort and had the necessary skills to solve the problem.

The second claim evaluated in this paper, that strong sampling is appropriate for describing human property induction, has little support in these data. Not only did few people follow the size principle given explicitly strong sampling, but ambiguous instructions almost invariably resulted in monotonicity. It may well be that strong sampling is the default assumption for other inductive tasks, but for the types of stimuli used in this experiment, adding premises within a range did not generally weaken property induction to a conclusion outside of that range.

Acknowledgments

This work was supported by a Brown University Graduate Fellowship. Many thanks Tom Griffiths, Steve Sloman and 4 anonymous reviewers for valuable input.

References

Anderson, J.R. (1990). *The adaptive character of thought:* Erlbaum.

Evans, J. St. B. T. & Over, D. E. (1996). *Rationality and reasoning.* Hove: Psychology Press.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition*: Oxford University Press.

Heit, E. (2001). What is the probability of the Bayesian model given the data? *Behavioral and Brain Sciences*, 24, 672-673. (commentary on Tenenbaum & Griffiths (2001)).

Kemp, C., & Tenenbaum, J. (2003). *Theory-based induction.* Paper presented at the Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin and Review, 10*(3), 517-532.

Osherson, D. N., Wilkie, O., Smith, E. E., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*(2), 185-200.

Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29*(6), 1141-1159.

Sanjana, N. E., & Tenenbaum, J. B. (2003). *Bayesian models of inductive generalization.* Paper presented at the Advance in Neural Information Processes 15.

Shafto, P., Kemp, C., Baraff, E., Coley, J. D., & Tenenbaum, J. (2005). *Context sensitive induction.* Paper presented at the Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*(4820), 1317-1323.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25*, 231-280.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

Sloman, S. A., & Lagnado, D. A. (2005). The problem of induction. In R. Morrison & K. Holyoak (Eds.), *Cambridge handbook of thinking and reasoning*. New York: Cambridge University Press.

Stanovich, K.E. & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645-726.

Tenenbaum, J. B. (2000). *Rules and similarity in concept learning.* Paper presented at the Advances in Neural Information Processing Systems 12.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity and Bayesian inference. *Behavioral and Brain Sciences, 24*, 629-640.